An Ex-Post Analysis of the Phenomenon of Wash Trading on NFTs

(Discussion Paper)

Gianluca Bonifazi[1,†], Francesco Cauteruccio[1,†], Enrico Corradini[1,†], Michele Marchetti[1,†], Domenico Ursino[1,*,†] and Luca Virgili[1,†]

[1]*DII, Polytechnic University of Marche*

Abstract

In recent years, many researchers have studied the phenomenon of wash trading on NFTs (Non Fungible Tokens) from an "ex-ante" perspective. The latter aims to identify and classify wash trading activities before or as they occur. In this paper, we propose an "ex-post" analysis of wash trading on NFTs. This perspective aims to analyze wash trading activities carried out in the past to see whether such illicit and risky activities brought significant profit to those who did them. To the best of our knowledge, this is the first paper in the literature to analyze wash trading on NFTs from an "ex-post", instead of an "ex-ante", perspective.

Keywords

Wash trading, Non Fungible Tokens, Blockchain, Correlation, Causality, Distance, Cryptoslam

1. Introduction

An NFT (Non-Fungible Token) is a digital content that represents a real-world object, e.g., artwork, music, game or collection. NFT technology presents many opportunities but also several risks [1, 2, 3]. One of the main risks involves market manipulation, achieved by artificially increasing the price of an NFT. This practice is known as wash trading and involves a series of sales and purchases on the same NFT, made by the same trader, with the aim of generating a false interest in it, so as to artificially increase its value. In the past literature, studies on NFTs are very diversified as they address various issues. Some of them are: *(i)* the analysis and characterization of NFT ecosystems [4, 5, 6], *(ii)* the study of the role of social media in the context of NFTs [7], *(iii)* the analysis of geopolitical risks and market factors in this scenario [8, 9, 10]. The phenomenon of wash trading on NFTs has also been extensively studied in the literature. However, all past studies have considered an "ex-ante" perspective of this

SEBD 2023: 31st Symposium on Advanced Database Systems, July 02–05, 2023, Galzignano Terme, Padua, Italy
*Corresponding author.
†These authors contributed equally.
g.bonifazi@univpm.it (G. Bonifazi); f.cauteruccio@univpm.it (F. Cauteruccio); e.corradini@pm.univpm.it (E. Corradini); m.marchetti@pm.univpm.it (M. Marchetti); d.ursino@univpm.it (D. Ursino); luca.virgili@univpm.it (L. Virgili)
0000-0002-1947-8667 (G. Bonifazi); 0000-0001-8400-1083 (F. Cauteruccio); 0000-0002-1140-4209 (E. Corradini); 0000-0003-3692-3600 (M. Marchetti); 0000-0003-1360-8499 (D. Ursino); 0000-0003-1509-783X (L. Virgili)

phenomenon. In other words, they sought to identify and classify wash trading activities before or as they occur [11, 12, 13, 14]. In contrast, in the literature, there are no "ex-post" analyses of wash trading, which aim to understand whether wash trading activities really succeeded in increasing the interest and value of the NFTs on which they were carried out. The results of this analysis could have so many applications. For example, since wash trading is an illicit, improper and potentially harmful practice, one can ask: Is it worth it? In the face of these costs and risks, what are the benefits? A study of Chainanalysis[1] points out that many wash trading activities on NFTs in the past have not brought in large profits. This report is certainly an interesting starting point for answering the previous questions. However, it is not the result of a "structured" analysis, that is a research taking into account the intrinsic meaning of the various features characterizing an NFT. Actually, a "structured" analysis could provide insights into the presence of correlations, causal relationships and other forms of relationships among the features of an NFT. If it turns out that such relationships do not exist, one could convince wash traders that continuing to engage in this illicit practice is not worth it. The "structured" and "ex-post" analysis of wash trading on NFTs is the main objective of this paper.

An NFT[2] is defined by several features (think, for instance, of sales volume, price, owner number, etc.). Since the values of these features vary over time, it will be necessary to manage the temporal dimension of this phenomenon. In addition, the various features may be related to each other by different relationships (think, for instance, of correlation, causality, distance, etc.). In the following, we will use the term "facet" to refer to such relationships because each of them allows us to view the phenomenon from a certain angle or dimension (a facet, in fact). These facets, when applied to the features of NFTs, allow us to perform a "structured" analysis of the wash trading phenomenon. However, such analysis would not be possible if we did not have a congruent model to represent NFTs. Therefore, in this paper, we will also define such a model.

The outline of this paper is as follows: in Section 2, we present the model and framework used to conduct our analysis. In Section 3, we describe some of the experiments we conducted along with a discussion of the results obtained. Finally, in Section 4, we draw our conclusion and sketch some possible future developments.

2. A framework to support our investigation on wash trading on NFTs

Before illustrating our framework for investigating wash trading on NFTs, it is necessary to describe the underlying model by which we represent a set of NFT collections.

An NFT collection is a set of NFTs of the same type created by the same author. Specifically, let $\mathcal{S} = \{C_1, C_2, \cdots, C_l\}$ be a set of NFT collections of interest. Our model assumes that all NFT collections are characterized by the same set $\mathcal{F} = \{f_1, f_2, \cdots, f_m\}$ of features. A generic NFT collection $C_i \in \mathcal{S}$ can be represented as:

$$C_i = \langle Name_i, \mathcal{F} \rangle$$

[1] https://blog.chainalysis.com/reports/2022-crypto-crime-report-preview-nft-wash-trading-money-laundering/

[2] In this paper, when we talk about an NFT, we mean a collection of homogeneous NFTs; therefore, in the following, we will use the terms "NFT" and "NFT collection" interchangeably.

$Name_i$ is the name of C_i and is unique. $\mathcal{F}$ consists of numerical features whose values may vary over time. Borrowing the concepts of time series analysis [15], we can model $\mathcal{F}$ as a multivariate series. Examples of features of $\mathcal{F}$ are sales price, number of sellers and number of owners.

As mentioned in the Introduction, the temporal dimension is crucial in our model. Given a time interval T of interest, we can think of modeling it as an ordered sequence of z time slices $T = T_1, \cdots, T_z$. For example, T could be a certain month, say April 2023. In this case, it could be represented by a succession of 30 time slices, one for each day. Our time model must have a mechanism for indexing the sequence of time slices so that a particular interval of contiguous time slices of T (e.g., the first decade of April 2023) can be selected. To this end, it provides the notation $T[x..y]$, $1 \leq x \leq y \leq z$, to represent the interval of contiguous time slices beginning at T_x and ending at T_y. If $x = y$, then it means that we want to select a single time slice. In this case, we will use the abbreviated notation T_x or $T[x]$, instead of $T[x..x]$. If $x = 1$ and $y = z$, then it means that we want to select the whole interval of interest. In this case, we will use the abbreviated notation T, instead of $T[1..z]$, to represent that interval.

The notation defined above for time intervals can be extended to the features of an NFT collection. Specifically, given the collection $C_i \in \mathcal{S}$, we indicate by $\mathcal{F}[x..y]$ the trend of the values of $\mathcal{F}$ assumed by C_i in the time interval $T[x..y]$. In turn, $f_j[x..y]$, $1 \leq j \leq m$, indicates the values of f_j assumed by C_i in the time interval $T[x..y]$. Clearly, $\mathcal{F}[x..y] = \{f_1[x..y], f_2[x..y], \cdots, f_m[x..y]\}$. We adopt the notation $\mathcal{F}[x]$, or $\mathcal{F}_x$, (resp., $f_j[x]$, or f_{j_x}) to indicate the values of $\mathcal{F}$ (resp., the value of f_j) assumed by C_i during the time slice T_x, i.e., $\mathcal{F}[x] = \mathcal{F}[x..x]$ (resp., $f_j[x] = f_j[x..x]$). Finally, we adopt the abbreviated notation $\mathcal{F}$ (resp., f_j) to indicate the trend of the values of $\mathcal{F}$ (resp., the value of f_j) assumed by C_i during the overall time interval T, i.e., $\mathcal{F} = \mathcal{F}[1..z]$ (resp., $f_j = f_j[1..z]$).

The facets that our framework uses to investigate the wash trading on NFTs phenomenon are three, namely correlation, causality and distance between features. They capture and allow us to analyze different aspects of the phenomenon. However, our framework is extensible and allows one or more of these facets to be removed and/or new facets to be added.

The first facet used in our framework is correlation. Correlation analysis is performed by means of two techniques. The first computes the correlation between two features by calculating the Pearson coefficient between the corresponding time series [16]. This technique is straightforward and can already provide interesting results. However, it does not consider cross-correlation, that is, the possibility that the two features are correlated but one of them lags behind the other. To account for this, we employ a second technique, namely cross-correlation. In order to define it, we must preliminarily standardize the time series associated with the features. To this end, given a value of a feature, we subtract from it the mean of the values of the feature and divide the result obtained by the corresponding standard deviation. After all features have been standardized, it is possible to calculate the cross-correlation between two features by considering the x^{th} element of one of them and the $(x + h)^{th}$ element of the other. The parameter h represents the lag we want to consider. In principle, h can take an integer value between 0 and $z - x$. In practice, h is generally low; for example, we considered values of h less than or equal to 2 in our experiments. Cross-correlation between two features allows us to understand whether the values of one feature can be used to predict the values of the other. Given two features, our framework calculates the cross-correlation for different values of h and selects the maximum value as the overall cross-correlation value between the two features. All details about this procedure can be found in [17]. Correlation is the facet of our framework indicating the similarity degree for each pair of features, regardless of the lag and whether or not that similarity is due to a cause-and-effect phenomenon (the latter aspect is taken into account by causality facet).

The second facet used in our framework is the causal relationship between two features. It expresses the fact that the values of one of them depend on the values of the other. To compute the causality degree between a pair of features f_j and f_k, we consider the test proposed by Granger [18]. It is based on the idea of comparing the ability to predict the values of f_k using all the information in the Universe $\mathcal{U}$ with the ability to predict the same values using all the information in $\mathcal{U}$ except the values of f_j. We adopt the notation $\mathcal{U}|f_j$ to indicate the latter. If discarding f_j reduces the predictive power of f_k, then it means that f_j has unique information about f_k, in which case we say that f_j Granger-causes f_k. Saying that f_j Granger-causes f_k implies that the knowledge of the values of f_j allows the prediction of the values of f_k. To apply the Granger-test, f_j and f_k must be stationary. The Augmented Dickey-Fuller test [19] can be adopted to perform such a verification. If one or more series are not stationary, a differentiation process must be applied to make them stationary. Full technical details about the Granger test and how to use it in our framework can be found in [17]. Causality represents a stronger relationship than correlation. If f_j causes f_k, we can also say that f_j is correlated to f_k, while the vice versa does not apply. However, just because two features can be correlated even though there is no causal relationship between them, correlation still remains a facet to be considered in our framework.

The third facet used in our framework is the distance between two features. Since, in our model, features represent time series, it is not possible to employ classical distance metrics, such as the Euclidean, Manhattan or Minkowski distance because they are incapable of determining whether two time series align but with some delay from each other. Therefore, it is necessary to consider a distance metrics that takes this factor into account. Among those available in the literature, we decided to use the Dynamic Time Warping (DTW) distance [20]. Full details on how the DTW distance can be applied in our framework can be found in [17]. Distance can be computed for each pair (f_j, f_k) of features, regardless of whether there is a correlation or causality relationship between them. From this point of view, distance is a universal "facet" because it can be employed in any circumstance. It is clear that, at equal distances, if two features are also related by a correlation or even causality relationship, there is a much stronger link between them that should be taken into account. And that is why it makes sense to consider the distance facet and, at the same time, the correlation and causality ones. This reasoning reinforces the idea that a multifacet approach to the wash trading on the NFTs phenomenon can provide more accurate results than an approach that considers only one form of relationship between features.

3. Experiments

We built the dataset for the experiments by extracting data from Cryptoslam[3], an aggregator of data on NFT sales involving the Ethereum, WAS and FLOW blockchains. Specifically, we downloaded data on all NFT sales made from the appearance of each NFT collection in the market until February 7, 2023. Then, from all NFT collections thus downloaded, we selected the top 2000 ones based on their USD sales volumes. In this way, we constructed a ranking from which to draw NFT collections for analysis. Because the amounts of money in the dataset were expressed some in USD and some in ETH, we downloaded the daily ETH/USD exchange rate from Yahoo Finance in order to uniform them. The complete dataset thus constructed consisted of 672,098 rows, with an average of 336 rows for each NFT collection.

At this point, we identified the features of NFT collections useful for studying the phenomenon of our interest. Specifically, we identified the following features: *(i) Name*: it denotes the name of the NFT collection to which all next features refer; *(ii) Date*: it represents the day to which all next features refer; *(iii) Sales*: it indicates the sales volume expressed in USD; *(iv) ETH*: it denotes the ETH/USD exchange rate; *(v) Floor Price*: it represents the minimum price of an NFT in the collection; *(vi) Active Wallets*: it indicates the number of active wallets; *(vii) Sales Txns*: it denotes the number of sale transactions; *(viii) Total Owners*: it represents the total number of owners; *(ix) Wash Sales*: it denotes the volume of wash sales expressed in USD; *(x) Wash Txns*: it indicates the number of transactions made to perform wash trading. The set $\mathcal{F}$ of the features of interest for our analysis is given by all these features except the first two.

After that, we homogenized the time slices. In fact, the various features had been surveyed at different cadences. To this end, we chose to adopt the daily cadence for all of them to ensure consistency.

The first test we performed focused on the correlation facet. In Section 2, we have seen that there are two correlation metrics, namely the Pearson coefficient and cross-correlation. We began this test by calculating, for each NFT collection, the Pearson coefficient related to each pair of features. Then, we averaged the corresponding values over all available NFT collections. The results obtained are shown in Figure 1. As can be seen from this figure, the correlation between Wash Sales and Sales is very low.

After that, we analyzed the cross-correlation between features. In particular, we considered three cases corresponding to a number h of lags equal to 0, 1 and 2. For each case, we computed the value of cross-correlation for each pair of features and for each NFT collection. Finally, we averaged the values thus obtained over all NFT collections. The results for $h = 0$ are shown in Figure 1, since, in this case, cross-correlation coincides with the Pearson coefficient. The results for $h = 2$ are shown in Figure 2. Those for $h = 1$ are not shown because of space limitations; they can be found in [17]. Again, looking at Figure 2 we can see that the correlation value between Wash Sales and Sales is very low. The same happened for $h = 1$.

The second test we conducted focused on causality. Specifically, we applied the Granger causality test to verify whether Wash Sales Granger-causes Sales. If this happened, we would have a clue that it is worthwhile to carry out wash trading activities on NFTs. If not, we would have a second clue, in addition to that provided by correlation analysis, that it is not

[3]http://criptoslam.io

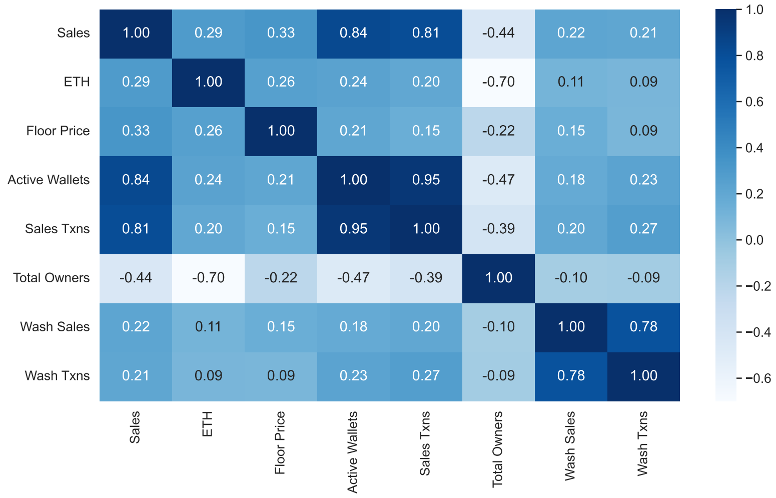

Figure 1: Values of the Pearson coefficient averaged on all available NFT collections

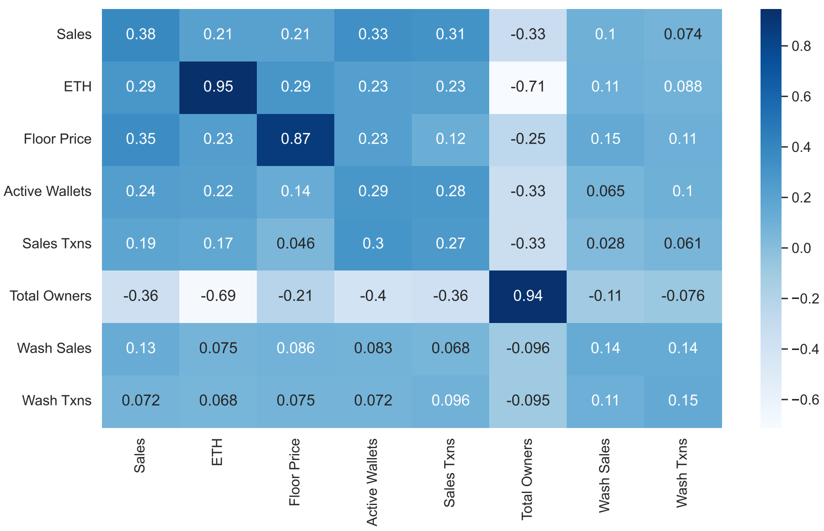

Figure 2: Values of the cross-correlations averaged on all available NFT collections for h=2

worthwhile. To carry out this verification, we considered the following null hypothesis: H_0: "Wash Sales does not Granger-cause Sales". For each NFT collection we calculated the p-value associated with the null hypothesis. Then, we computed for how many NFT collections the p-value was less than 0.05. We performed this calculation both using a linear autoregressive model (VAR) and adopting two deep learning models, namely MultiLayer Perceptron (MLP) and Long Short-Term Memory (LSTM). Finally, we considered different values of p (representing the maximum number of lagged observations), namely $p = 1$, $p = 2$ and $p = 3$. The results obtained are shown in Table 1.

Model	$p = 1$	$p = 2$	$p = 3$
VAR	4.25%	4.03%	4.01%
MLP	1.04%	1.04%	0.52%
LSTM	2.13%	1.54%	1.53%

Table 1
Percentage of the NFT collections for which the null hypothesis H_0 can be rejected

From the analysis of this table, we can deduce that the null hypothesis is almost always confirmed with all the three models considered. This implies that we have a second clue that wash trading on NFTs is not worth doing. However, before drawing any firm conclusion on this, we thought it was appropriate to consider the third facet, namely distance.

The third test we conducted focused on distance. Specifically, we calculated the value of DTW for each pair of features and for each NFT collection in the dataset. Then, we averaged the DTW values thus obtained over all the NFT collections. The final results are shown in Figure 3. From the analysis of this figure, we can observe that the DTW value for the pair (Wash Sales, Sales) is not only high but even the highest among all pairs of features.

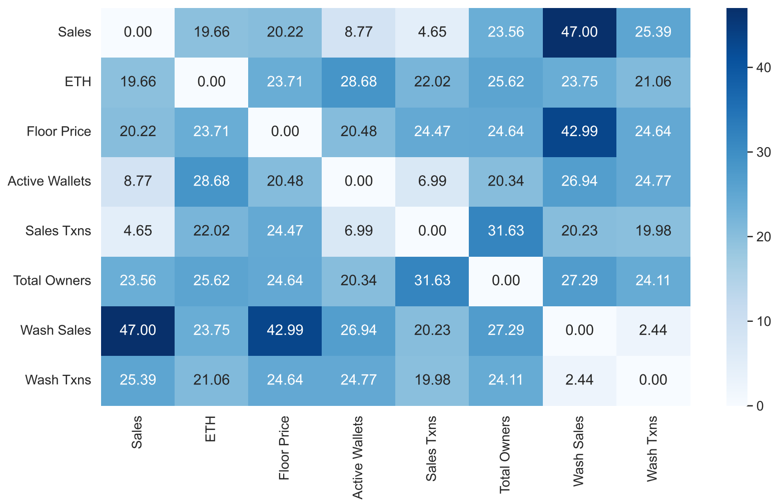

Figure 3: Distance values averaged on all available NFT collections

All these experiments allow us to conclude that our framework is capable of supporting NFT sales analysis and, in particular, wash trading investigation. In fact, it obtained interesting and concordant results for all the three facets considered. First the Pearson coefficient and cross-correlation values, then the causality values and finally the distance values say us that there is no significant relationships between Wash Sales and Sales. This allows us to conclude that it is very unlikely that there is an impact of wash trading activities on the economic values of an NFT collection.

4. Conclusion

In this paper, we have proposed an "ex-post" analysis of the wash trading on NFTs phenomenon. While in the literature there are several papers that present "ex-ante" analyses of this phenomenon, to the best of our knowledge this is the first paper that provides an "ex-post" perspective. The latter had been considered in the past by blogs and magazines that, however, had only presented simple statistics on this phenomenon. Instead, we have proposed a framework, with a well-defined underlying data model, capable of supporting this type of analysis. A key concept in our model is the facet one, which is a novelty introduced by our very framework. At the end of our analysis, we were able to conclude that it is not worthwhile to engage in such an illicit and risky practice as wash trading on NFTs.

The results obtained in this paper are not to be considered as an ending point, but rather as a starting point for further research on the wash trading on NFTs phenomenon. An initial development could involve analyzing the relationships between other features of NFTs, such as rarity and asset type, and wash trading activity. Indeed, some types of NFTs might be more prone to wash trading than others. Categorizing NFTs based on their features may allow a fine-grained study on the phenomenon of wash trading for the various types of NFTs. This could help determine the extent to which a new NFT collection might be subject to speculation. A next development could regard the application of machine learning algorithms to extract knowledge patterns related to the features involved in the wash trading phenomenon. Finally, we would like to study what factors underlie the pricing of NFTs, focusing in particular on the community that supports an NFT project. Specifically, we plan to analyze the social media ecosystem around an NFT project and characterize the behavior of users who participate in it, as well as the characteristics of the content they produce.

References

[1] K. Wilson, A. Karg, H. Ghaderi, Prospecting non-fungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity, Business Horizons 65 (2022) 657–670. Elsevier.

[2] D. Wang, S. Wu, Z. Lin, L. Wu, X. Yuan, Y. Zhou, H. Wang, K. Ren, Towards a first step to understand flash loan and its applications in defi ecosystem, in: Ninth International Workshop on Security in Blockchain and Cloud Computing (SBC '21), Hong Kong, China, 2021, pp. 23–28. ACM.

[3] L. Ante, Non-fungible token (NFT) markets on the Ethereum blockchain: Temporal development, cointegration and interrelations, Economics of Innovation and New Technology (2022) 1–19. Taylor & Francis.

[4] M. Nadini, L. Alessandretti, F. D. Giacinto, M. Martino, L. M. Aiello, A. Baronchelli, Mapping the NFT revolution: market trends, trade networks, and visual features, Scientific reports 11 (2021) 20902. Nature Publishing Group UK London.

[5] K. Vasan, M. Janosov, A. L. Barabási, Quantifying NFT-driven networks in crypto art, Scientific reports 12 (2022) 1–11. Springer.

[6] B. White, A. Mahanti, K. Passi, Characterizing the OpenSea NFT marketplace, in: ACM Web Conference 2022 (TheWebConf'22)), Virtual Event, 2022, pp. 488–496. ACM.

[7] S. Casale-Brunet, M. Zichichi, L. Hutchinson, M. Mattavelli, S. Ferretti, The impact of NFT profile pictures within social network communities, in: ACM Conference on Information Technology for Social Good (GOODIT '22), Limassol, Cyprus, 2022, pp. 283–291. ACM.

[8] C. Urom, G. Ndubuisi, K. Guesmi, Dynamic dependence and predictability between volume and return of Non-Fungible Tokens (NFTs): The roles of market factors and geopolitical risks, Finance Research Letters 50 (2022) 103188. Elsevier.

[9] D. Costa, L. L. Cava, A. Tagarelli, Show me your NFT and I tell you how it will perform: Multimodal

representation learning for NFT selling price prediction, in: Proceedings of the ACM Web Conference 2023, Austin, Texas, USA, 2023, pp. 1875–1885.
[10] L. L. Cava, D. Costa, A. Tagarelli, Visually Wired NFTs: Exploring the Role of Inspiration in Non-Fungible Tokens, arXiv preprint arXiv:2303.17031 (2023).
[11] S. Serneels, Detecting wash trading for nonfungible tokens, Finance Research Letters (2022) 103374. Elsevier.
[12] V. V. Wachter, J. R. Jensen, F. Regner, O. Ross, NFT Wash Trading: Quantifying suspicious behaviour in NFT markets, arXiv preprint arXiv:2202.03866 (2022).
[13] D. Weber, Exploring markets: Non-Fungible Token, Technical Report, IU Discussion Papers-Business & Management, 2022.
[14] N. Tahmasbi, A. Fuchsberger, Non-fungible Tokens-Exploring Suspicious Washtrader Communities in NFT Networks, in: International Conference on Information System (ICIS'22), 1467, Copenhagen, Denmark, 2022.
[15] G. Reinsel, Elements of multivariate time series analysis, 2003. Springer Science & Business Media.
[16] K. Pearson, Note on Regression and Inheritance in the Case of Two Parents, Proceedings of the Royal Society of London 58 (1895) 240–242. The Royal Society.
[17] G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, D. Montella, S. Scarponi, D. Ursino, L. Virgili, Performing Wash Trading on NFTs: is the Game Worth the Candle?, Big Data and Cognitive Computing 7 (2023). MDPI.
[18] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: Journal of the Econometric Society (1969) 424–438. JSTOR.
[19] D. Dickey, W. Fuller, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74 (1979) 427–431. Taylor & Francis.
[20] M. Müller, Dynamic time warping, Information Retrieval for Music and Motion (2007) 69–84. Springer.